RESEARCH ARTICLE

Unidata's Common Data Model mapping to the ISO 19123 Data Model

Stefano Nativi · John Caron · Ben Domenico ·
Lorenzo Bigagli

Abstract Access to real-time distributed Earth and Space Science (ESS) information is essential for enabling critical Decision Support Systems (DSS). Thus, data model interoperability between the ESS and DSS communities is a decisive achievement for enabling cyber-infrastructure which aims to serve important societal benefit areas. The ESS community is characterized by a certain heterogeneity, as far as data models are concerned. Recent spatial data infrastructures implement international standards for the data model in order to achieve interoperability and extensibility. This paper presents well-accepted ESS data models, introducing a unified data model called the Common Data Model (CDM). CDM mapping into the corresponding elements of the international standard coverage data model of ISO 19123 is presented and discussed at the abstract level. The mapping of CDM scientific data types to the ISO coverage model is a first step toward interoperability of data systems. This mapping will provide the abstract framework that can be used to unify subsequent efforts to define appropriate conventions along with explicit agreed-upon encoding forms for each data type. As a valuable case in point, the content mapping rules for CDM grid data are discussed addressing a significant example.

Keywords Common data model · Coverage · Data model · Fluid earth sciences · ISO 19123 · Model mapping · Scientific data types

S. Nativi (✉) · L. Bigagli
IMAA, Italian National Research Council,
C.da S.Loja–Zona Industriale,
Tito Scalo, Italy
e-mail: nativi@imaa.cnr.it
URL: www.imaa.cnr.it

L. Bigagli
e-mail: bigagli@imaa.cnr.it
URL: www.imaa.cnr.it

S. Nativi · L. Bigagli
University of Florence,
Florence, Italy

J. Caron · B. Domenico
Unidata Program Center, UCAR Office of Programs,
P.O. Box 3000, Boulder, CO 80307-3000, USA

J. Caron
e-mail: caron@ucar.edu
URL: www.unidata.ucar.edu

B. Domenico
e-mail: domenico@ucar.edu
URL: www.unidata.ucar.edu

Introduction

Several interoperability projects, experiments and test-beds have implemented and tested a geoinformation standards-based interface to the wealth of Earth science datasets that are currently available in formats well-accepted in their own communities of practice. Prominent examples are: netCDF (NetCDF: http://www.unidata.ucar.edu/software/netcdf/), HDF (HDF group: http://hdf.ncsa.uiuc.edu/index.html) and GRIB (GRIB: http://www.wmo.ch/pages/prog/www/WMOCodes/GRIB.html) form. These are often served (directly or indirectly) via the OPeNDAP (OPeNDAP: http://www.opendap.org/) client–server protocol. A valuable case in point is the OGC GALEON (Geo-Interface to Air, Land, Environment, Oceans NetCDF) interoperability experiment (http://galeon-wcs.jot.com/WikiHome).

Indeed, there is a clear need to establish a solid interoperability framework between these two important communities. In fact, the netCDF community itself is really multi-disciplinary—spanning several realms, including atmospheric sciences, oceanography, hydrology, etc. Like-

wise the Geospatial Information (Geoinformation) community has expanded well beyond its initial roots in the traditional GIS (Geographic Information Systems) community and is becoming more and more important in the present internet web services era. In fact, access to real-time distributed geospatial information is essential for enabling critical decision support systems. Therefore, netCDF community standards, ISO/OGC standards and the related crosswalk solutions establish a unique framework for supporting the Information Society, facilitating the provision of real time geosciences data to decision support systems.

This document describes the relationships between the underlying data models used in these communities. In particular it describes the Unidata Common Data Model (CDM) (Common Data Model: http://www.unidata.ucar.edu/software/netcdf/CDM/index.html) and maps it into corresponding elements of the international standard coverage data model of ISO 19123 (ISO/FDIS 19123 Geographic information—Schema for coverage geometry and functions, ISO/FDIS 19123 2005).

Conceptual model mapping for implementing full Interoperability

In a general framework, interoperability encompasses three different aspects (European Commission 2006):

- Semantic, the objective of which is ensuring the precise meaning of exchanged information is understandable by any application involved;
- Technical, which is concerned with the technical issues of linking up computer systems, the definition of open interfaces, data formats and protocols;
- Organizational, which deals with modeling organizational processes, aligning information architectures with organizational goals, and helping these processes to co-operate.

Semantic interoperability means enabling different agents, services, and applications to exchange information, data and knowledge in a meaningful way, on and off the Web (W3C 2005). Therefore, semantic interoperability is a necessary component in achieving full interoperability since it is concerned with ensuring that the precise meaning of exchanged information is understandable by other parties (IADBC 2005).

The ESS domain cannot be expressed adequately with a taxonomy, or with a thesaurus which models term relationships, as opposed to concept relationships. Therefore, conceptual models (i.e. models that define concepts of a universe of discourse) have been used in order to model the portion of the domain for which a database/file system provides data or for which an infrastructure provides

services. The Unified Modeling Language (UML) is the paradigmatic modeling language (Booch et al. 1998) used by domain experts.

These conceptual models require human semantic interpretation; hence, conceptual or abstract interoperability is a sine qua non for heterogeneous data models semantic interoperability. Even if semantic understanding is about more than mapping high-level concepts, it is important to define mappings between concepts within the data models, which requires content analysis. The present work focuses on this semantic interoperability aspect.

Abstract level mapping is important to constrain the possible mapping rules to be defined at the logical and, then, physical levels. In fact, given two data models, several logical mappings are meaningful. This situation clearly emerges when different disciplinary groups implement standard discovery and access services, such as the OGC WCS (Web Coverage Service) or OGC CSW (Catalog Service for Web). They map a community standard data model (e.g. CF-netCDF) to an international standard model (e.g. ISO 19123 or 19115) and produce different logical, and consequentially physical, mappings which are acceptable in the absence of a "commonly recognized" data mapping to the abstract model.

This work proposes the abstract mapping model for the CDM scientific data types and the ISO Coverage types, focusing on the abstract and content mapping between the CDM Grid and ISO Discrete Grid Coverage types. A set of mapping rules, expressed in natural language, are introduced which should be applied as constraints for implementing the logical and physical mappings. Implementation may be realized using different technological frameworks, such as: XML schemas and XSLT, object-oriented classes and mapping operations. Finally, an example of the result obtained applying the proposed mapping rules is reported for a complex CDM Grid dataset.

Data models

In a philosophical sense, a data model is a way of thinking about scientific data by applying a data model theory. It is an abstraction that describes how datasets are represented and used. In fact, an abstract model is a formal description of how data may be structured and used. Some of these abstract models have been incorporated into systems for storing and accessing scientific data. Different modelers may well produce different models of the same domain; where the data models differ significantly, it can be challenging to make the data systems interoperate with one another, which in turn, can stifle interdisciplinary research by hindering integrated analysis and display of

multiple datasets from different domains. Data modeling involves a progression from abstract model to logical and physical schemas.

In computer terms, a data model can be thought of as equivalent to an abstract object model in Object Oriented Programming in that an abstract data model describes data objects and what methods can be used on them.

Data model implementations

An abstract data model can be implemented in several forms (i.e. data model logical and physical schemas); it depends on the data manipulation technology used. For example:

- Object Oriented technology: an API is the interface to the data model for a specific programming language;
- Semi structured data technology: XML schema and file format are the artifacts that specify how to persist the objects in the data model.
- Structured data technology: SQL Data Definition Language schema and file format are the artifacts to specify the way to persist the objects in the data model.
- Service Oriented technology: a data access protocol plays the role of both API and file format for data exchanges over a network. Agreed upon data models are needed to understand the datasets that are transferred via such services.

The abstract data model, on the other hand, removes the details of any particular API and persistence format in which datasets are actually stored.

An earth sciences common data model (CDM)

The Common Data Model (CDM) is an effort to fuse the best characteristics of existing Earth science data models into one which is more powerful than each of the others, but maintains the fundamental simplicity and ease of use of the original netCDF. As depicted in Fig. 1, the resulting CDM consists of several layers. The top layer provides interfaces to a set of scientific data types. The middle layer provides access to coordinate system information and, at the bottom is the data access layer.

Existing earth science data models

Referring to Fig. 1, at its lowest (data access) layer, the CDM combines the most valuable features of three widely used Earth science data models: netCDF, HDF (HDF group: http://hdf.ncsa.uiuc.edu/index.html), and OPeNDAP (OPeNDAP: http://www.opendap.org/). The underlying data models are described in the following sections.
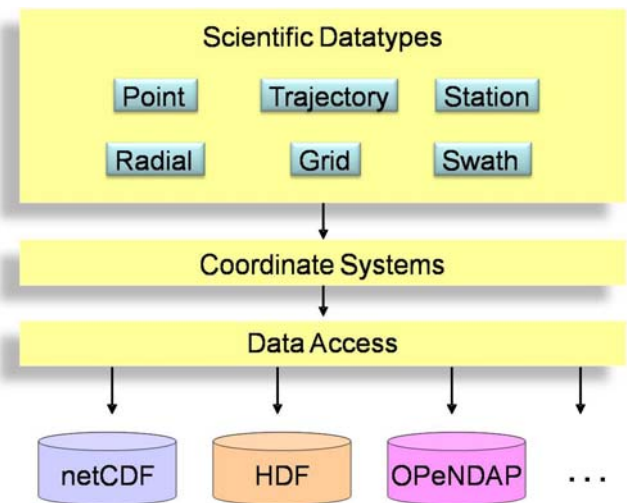

Fig. 1 Common data model layers

NetCDF-3

The netCDF-3 data model is fairly simple; it is shown in the Unified Modeling Language (UML) diagram of Fig. 2. A dataset has dimensions, variables, and attributes. Attributes can be global or apply to individual variables. There is a very limited set of low level data types.

Semantic metadata via CF-conventions for netCDF-3

In order to introduce more specific semantic elements (i.e. metadata) which are required by different communities to fully describe their datasets, the netCDF data model has been extended by adding a set of conventions. One of the most popular conventions is the Climate and Forecasting
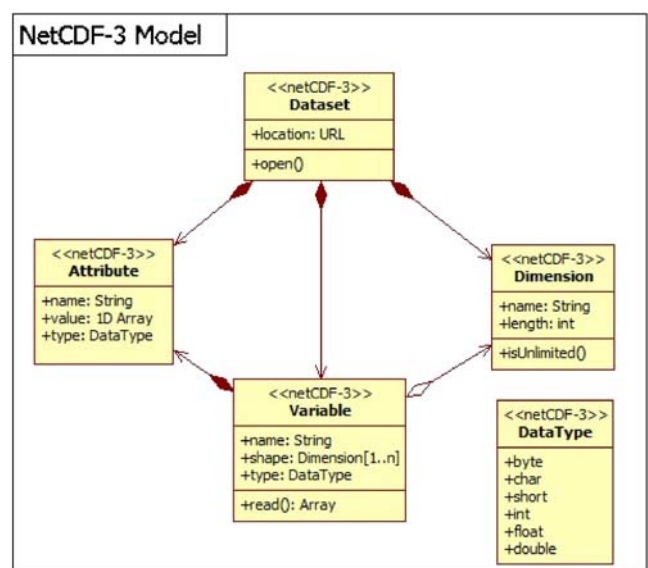

Fig. 2 netCDF-3 data model

metadata convention (CF) (CF: CF: Climate and Forecast Conventions: htto://cf-pcmdi.llnl.gov/; CF Standard Name Table: http://cf-pcmdi.llnl.gov/documents/cf-standard-names/; BADC Datasets: CF conventions: http://badc.nerc.

ac.uk/help/formats/netcdf/index_cf.html). Figure 3 depicts the CF-netCDF data model. CF conventions are quite loose, to maximize backward compatibility with the earlier COARDS (Cooperative Ocean/Atmosphere Research Data
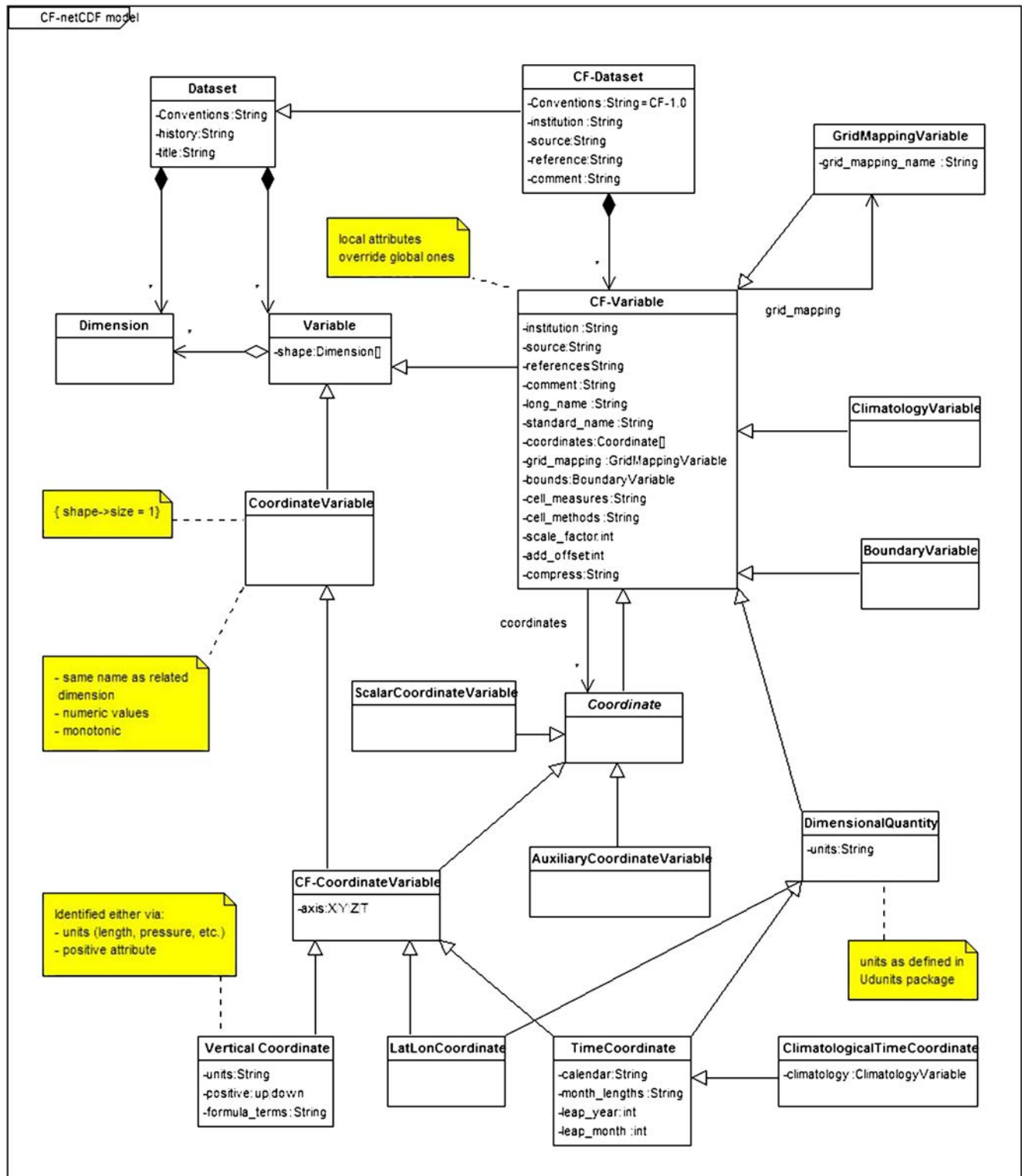

Fig. 3 CF-netCDF data model

Service) conventions. Besides, support for precise geo-location is scarce. For example, CF metadata conventions assume that "Latitude, longitude, and time are defined by internationally recognized standards, and hence, identifying the coordinates of these types is sufficient to locate data values uniquely with respect to time and a point on the earth's surface". On the other hand, the CF metadata model is very flexible and, consequently, more complex. Figure 3 diagram depicts CF conventions and their relationship with netCDF concepts, in UML.

The CDM data model includes most of the CF extensions, adopting their structure and semantics; a valuable case in point is the coordinate system entity. The few remaining CF metadata entities (e.g. coordinate axis units) can be easily implemented by profiling the CDM data model, working out a CF-CDM profile.

OPeNDAP

The OPeNDAP data model has many things in common with netCDF. But t has a richer set of low level data types and includes structures, sequences and grids. Figure 4 depicts the OPeNDAP UML schema.

HDF-5

HDF-5 has a much richer set of low level data types and includes the key feature of a group of variables. As with OPeNDAP, HDF-5 includes structures. Its schema is shown in Fig. 5.

All these abstract data models are fairly simple. In fact, the modeled domain is a generic, and optionally complex, dataset, describing its structure. The dataset semantic or context is not described, providing the user with a generic "attribute" element to capture and formalize more knowledge. The result is a flexible data model which can be used to model both simple and complex structured data, throughout almost all the Earth science disciplines.

On the other hand, these characteristics of generality and flexibility present challenges when it comes to semantic interoperability with other data systems. As noted above for netCDF, a partial answer to this problem was provided by the specification of netCDF conventions (NetCDF conventions: http://www.unidata.ucar.edu/software/netcdf/conventions.html); in fact, these supplementary specifications provide extensions to the underlying netCDF data model.

CDM: access layer

At the data access level, the CDM maintains as much as possible of the elegance of the netCDF-3 interface, but
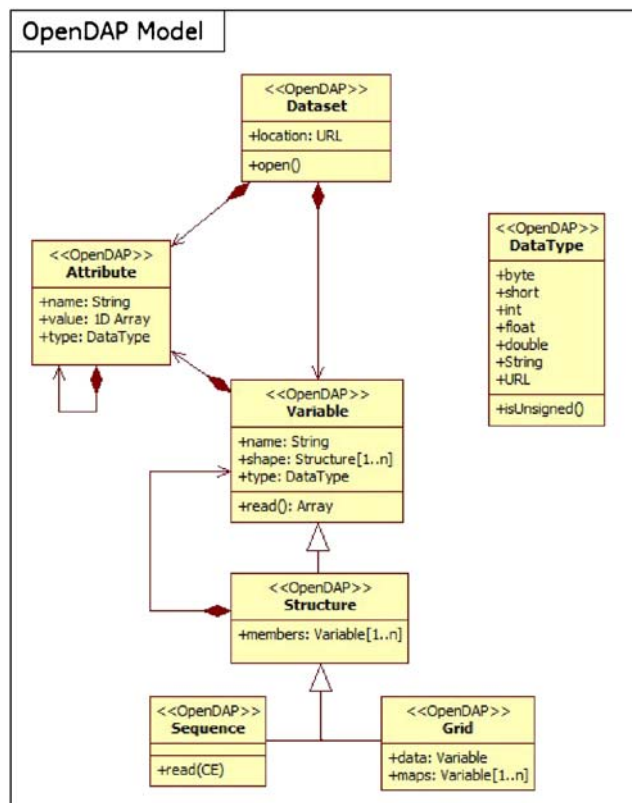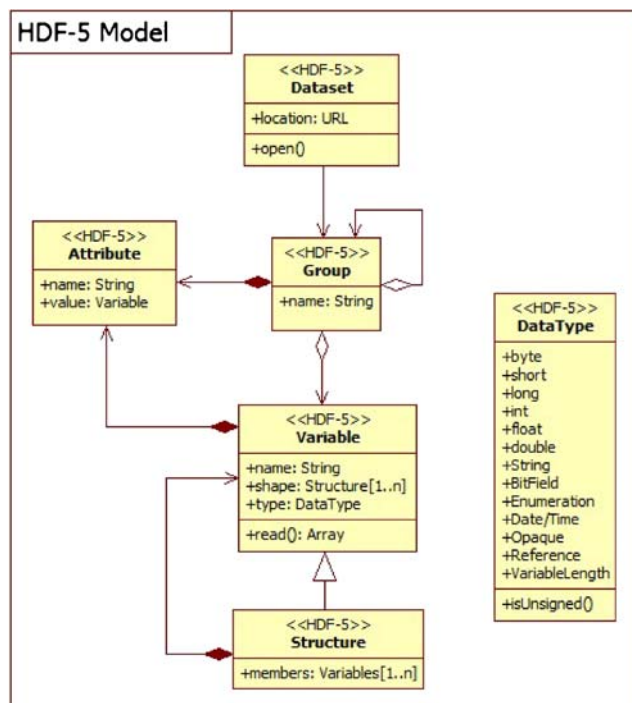

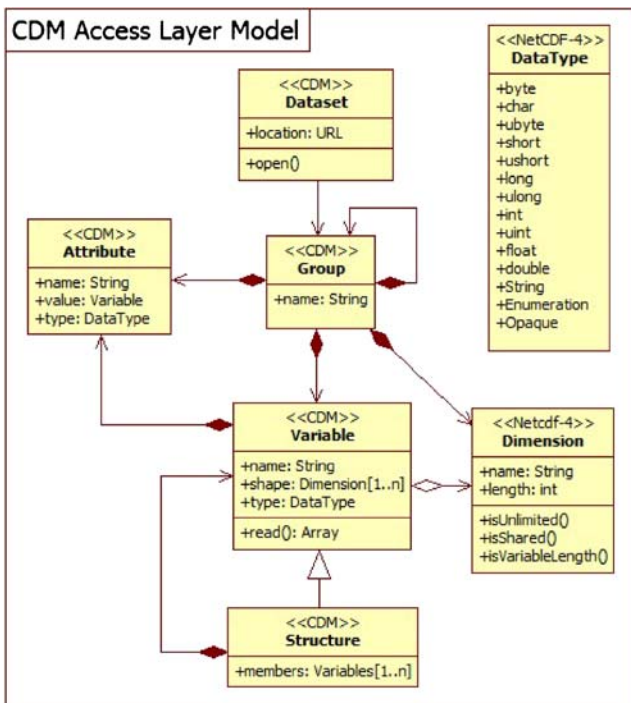**Fig. 4** OPeNDAP (DAP-2) data model

Fig. 5 HDF-5 data model

Fig. 6 UNIDATA common data model (data access layer)

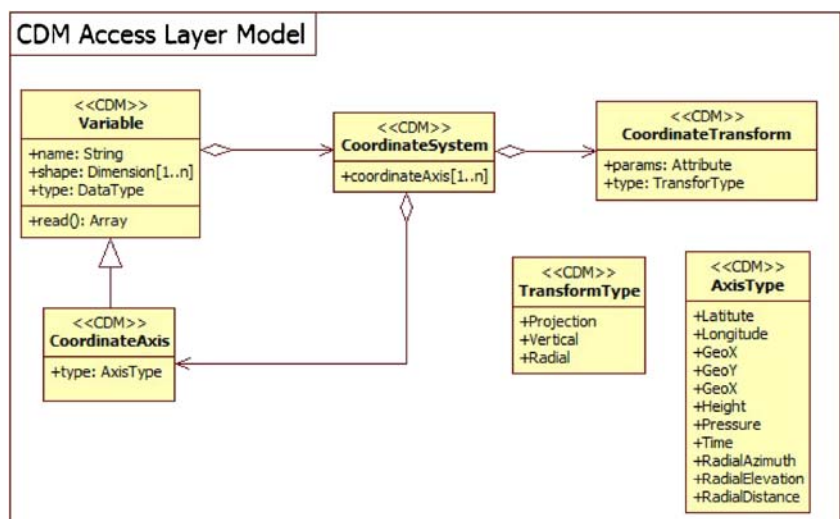adds important features from OPeNDAP and HDF, most notably:

- More low level data types—including "string"
- Structures
- Groups

Figure 6 shows the resulting object schema in UML notation.

CDM: coordinate systems layer

The netCDF, OPeNDAP, HDF data models do not have integrated coordinate systems, so georeferencing is not a

part of the API. As a consequence, the coordinate system information is inferred. In the best case, the files conform to a set of established conventions [e.g. CF-1, COARDS, etc (NetCDF conventions: http://www.unidata.ucar.edu/soft ware/netcdf/conventions.html)]. In contrast, for GRIB, HDF-EOS, and other specialized formats, the coordinate system specifications are built in.

In the CDM, the coordinate system information must be handled in a general way. The approach is shown in the diagram depicted in Fig. 7.

CDM: scientific data types

The top layer of the CDM carries the semantics in terms of a set of "Scientific Data Types." The distinction among the types is based on how the data points are connected. At this time, it is still in flux as the specific APIs are still evolving, but, in its current form, it is based on datasets types familiar to the netCDF community. In concept, the design scales to large, multi-file collections and will eventually support "specialized queries" such as those in space and time that are not part of the underlying netCDF. And more to the point in this article, it is intended to be used in the creation of standard netCDF file encoding conventions.

With the detailed definitions still evolving to some extent, the types are can cast into three main groups:

- Gridded Data
    - Structured
    - Swath
    - Unstructured

- Point Observation
    - Unconnected
    - Station observations/Time series

Fig. 7 UNIDATA common data model *Coordinate System*

Fig. 8 Illustrations of gridded data types
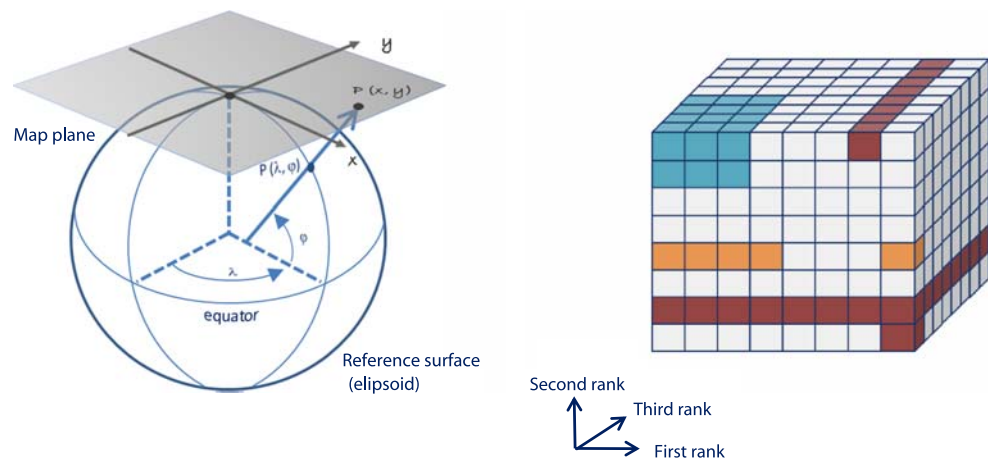
○ Trajectory
○ Profile

• Radial

Gridded data

Gridded data are specified in Cartesian coordinate systems with three spatial dimensions and time. The coordinates of the points are not specified explicitly but are implicitly determined by an algorithm. All dimensions are connected in the sense that neighbors in index space are neighbors in coordinate space. In a simple example, points are spaced in equal increments in longitude, latitude, height and time. Of course the relationships among the points in a swath or unstructured grid dataset are more complicated. For example as the coordinates of points in a swath image from a satellite depend on satellite navigation parameters, scanning rates, and so forth. Figure 8 shows examples of gridded data.

Point observations

For point observations, the coordinates are specified more explicitly. In the most general case, the coordinates of the point bear no relationship to one another. For example, for collections of lightning strike data, the coordinates are completely unconnected, so the spatial coordinates and time for each strike have to be specified explicitly in the dataset. On the other hand, station observations are taken at sets of points that remain fixed in space so the spatial coordinated can be specified in a table and the individual observations constitute a time series at each station. For a trajectory, the coordinates have to be specified explicitly but the points are an ordered set in time. A vertical profile is similar to a station observation dataset except that the vertical coordinate changes with each subsequent observation. Figure 9 shows examples of point observations.

Radial datasets

Radial datasets are a common data type in the atmospheric sciences associated with ground-based radar observations. As with gridded datasets, all dimensions are connected so neighbors in index space are neighbors in coordinate space. However, the spatial relationships are specified in polar coordinates of distance, azimuth, and elevation. Figure 10 depicts examples of radial datasets.

A general discussion of the CDM scientific data types is present on the Unidata's Common Data Model and THREDDS Data Server web pages[1]

The current version of the CF metadata conventions for netCDF have been developed primarily for the structured grid scientific data type, so this paper will confine itself to defining the mapping of the structured data type to the coverage data model used in international standards. As the equivalent conventions evolve for the other CDM scientific data types, equivalent mappings will be developed and published. In particular work is underway for defining a set of netCDF conventions for station observations datasets.
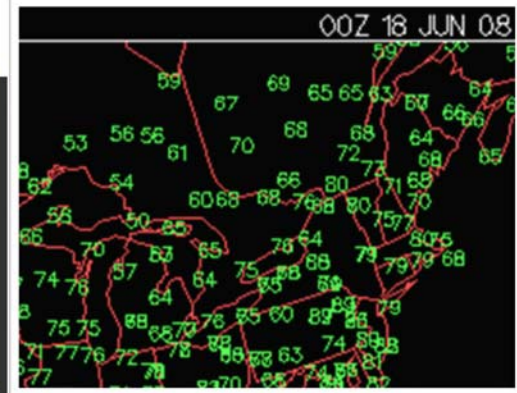
Interoperability via international standards

The technological components of the CDM have evolved as de facto standards over the last couple decades in the communities they serve. In particular, the atmospheric science and oceanography communities (sometimes referred to as the Fluid Earth Sciences or FES) have taken advantage of netCDF, HDF, and OPeNDAP. During the same period, other disciplines (notably solid Earth, hydrology, and human impacts) have employed

---

[1] http://www.unidata.ucar.edu/projects/THREDDS/CDM/CDM-TDS.htm

Fig. 9 Illustrations of point observation data types (source: Plymouth State University Weather Center - http://vortex.plymouth.edu)

Geographic Information Systems (GIS) technologies where the data models are quite different from those of the CDM. One approach to achieving interoperability between the data systems in these communities is to employ evolving international standards, especially those promulgated by the OGC (Open Geospatial Consortium) (OGC: Open Geospatial Consortium: http://www.opengeospatial.org/) and the ISO (International Organization for Standardization) technical committee on Geographic information/Geomatics (TC 211) (ISO TC211: Technical Committee on Geographic Information/Geomatics: http://www.isotc211.org/). ISO TC211 has developed a very elaborate and complete set of abstract data models for geospatial information.

Indeed the FES community deals with geospatial phenomena. FES data capture and represent discrete and continuous real world phenomena. Discrete phenomena are recognizable objects that have relatively well-defined boundaries or spatial extent (e.g. measurement stations). While, continuous phenomena vary over space and have no specific extent (e.g. temperature field); continuous phenomenon value is only meaningful at a particular position in space and time. ISO TC211 introduced two fundamental concepts to map both discrete and continuous real world phenomena: features and coverages. *A coverage is a feature that has multiple values for each attribute type, where each direct position within the* *geometric representation of the feature has a single value for each attribute type* (ISO/FDIS 19123 2005).

Historically, geospatial information has been managed in terms of two fundamental types called vector and raster data.

Vector data deals with discrete phenomena, each of which is conceived of as a feature (ISO/FDIS 19123 2005). The spatial characteristics of a discrete real-world phenomenon are represented by a set of one or more geometric primitives (e.g. points, curves, surfaces or solids) (ISO/FDIS 19123 2005). While the other phenomenon characteristics are treated as feature attributes. Generally, a single feature is associated with a single set of attribute values (ISO/FDIS 19123 2005). ISO 19107 (ISO/IS 19111 2003) provides a schema for describing features in terms of geometric and topological primitives.

Raster data deals with real-world phenomena that vary continuously over space (ISO/FDIS 19123 2005). It contains a set of values, each associated with one of the elements in a regular array of points or cells. Raster data is a commonly used example of Coverage. In fact, the coverage concept generalizes and extends the raster structure type by referring to any data representation that assigns values directly to spatial position. A coverage associates a position within a spatial/temporal domain to a value of a defined data type. It realizes a function from a spatial/temporal domain to an attribute domain (the co-domain) (ISO/FDIS 19123 2005).

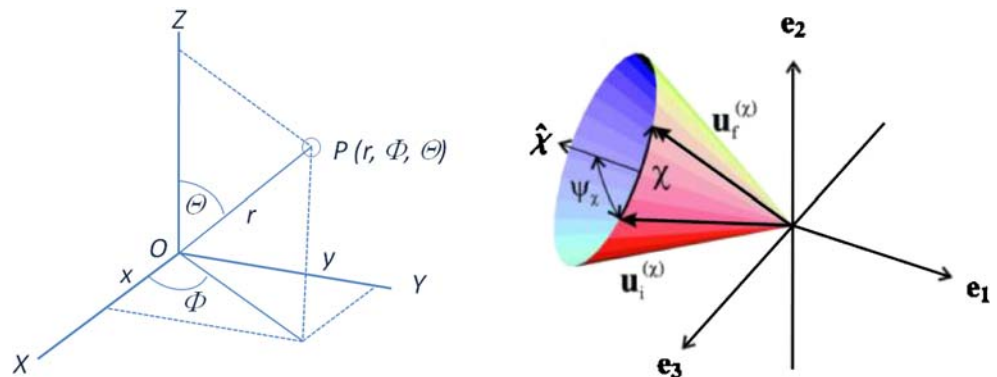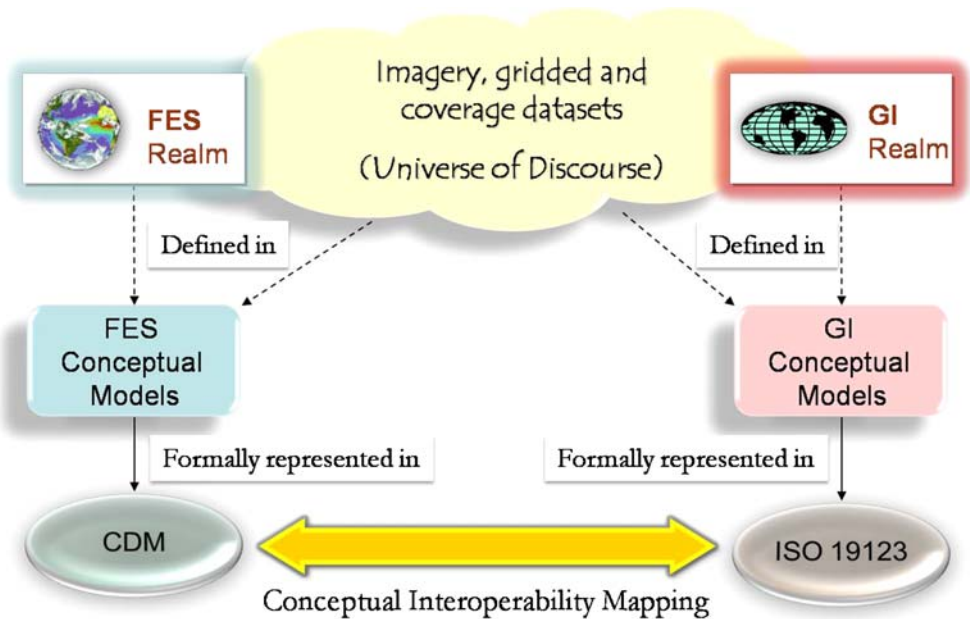**Fig. 10** Illustrations of radial data types

Just as the concepts of discrete and continuous phenomena are not mutually exclusive, their representations as discrete features or coverages are not mutually exclusive. The same phenomenon may be represented as either a discrete feature or a coverage (ISO/FDIS 19123 2005). However, Coverages are the prevailing data structures in FES community.

ISO has defined the ISO 19123 standard specification for imagery gridded and coverage data models. Therefore, the mapping between the CDM data model and the ISO 19123 is a key foundation component for establishing interoperability between the data systems in the realms of FES and

Geospatial information (GI) technologies. Figure 11 depicts this general framework.

ISO 19123 data model

The ISO definition of a coverage is: ...*a feature that associates positions within a bounded space (its domain) to feature attribute values (its range). In other words, it is both a feature and a function. Examples include a raster image, a polygon overlay or a digital elevation matrix* (ISO/FDIS 19123 2005). Figure 12 shows the coverage types introduced by ISO 19123.

Fig. 12 ISO 19123 coverage types
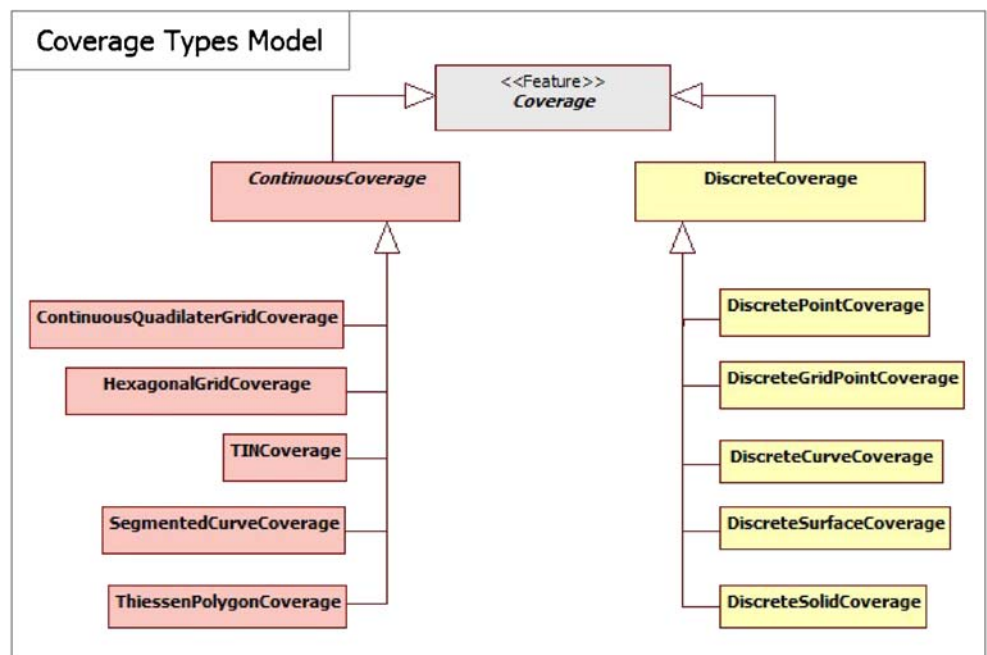
As far as the general geo-information framework is concerned, a coverage is a "feature" sub-type. In fact, a coverage is still an abstraction of the real world (i.e. an observation feature of interest) that has a spatial/temporal object as an attribute. This point is important to conceive a general framework for Earth phenomena observation.

The ContinuousCoverage type is the subclass of Coverage that returns a distinct record of feature attribute values for any direct position within its domain. The domain of a Discrete-Coverage consists of a collection of geometric objects or points in space. DiscreteCoverages are divided into subclasses on the basis of the type of geometric object in the spatial domain.

ISO abstract data models employ the language of mathematical function in the sense that the domain can be thought of as the set of values of independent variables defining positions in 3-dimensional space and time while the range is the set of values that the function takes on at those points in space.

Mapping CDM scientific data types to ISO 19123 coverage types

The ISO 19123 abstract model may be used to model the entire suite of CDM Scientific data types. Table 1 shows a high-level abstract mapping for the CDM Scientific data types to the corresponding ISO coverage type.

ISO Coverage consists of a set of domain objects (i.e. DomainObjects) which characterize the coverage domain. These objects represent an element of the coverage domain that may include any combination of geometric object [i.e. the Object types defined in the ISO 19107 standard (ISO/IS 19107 2003)], temporal geometric primitives [i.e. the temporal GeometricPrimitives defined in the ISO 19108 standard (ISO/IS 19108 2002)], or spatial and temporal objects defined in other standards, such as the GridPoint (defined in the same ISO 19123 standard). Figure 13 depicts this association.

Referring to Fig 13, it is noteworthy that a coverage is characterized by a general domain consisting of objects (i.e. DomainObject) which are composed of spatial and temporal objects or primitives. The domain nature is defined by the associated Coordinate Reference System [i.e. the CRS data types defined in the ISO 19111 standard (ISO/IS 19111 2003)], which is mandatory. Often, in the FES realm, the Coordinate Reference System is a Spatial&Temporal compound system.

CDM station time series, swath and radial scientific data types

Sometimes, for efficiency reasons the coverage function domain may be split up; a valuable example is a station time series. It may be modeled as a coverage characterized by a spatial domain generating a coverage domain element for each station location; then, a coverage attribute value record is associated for each domain element. Records values are parameterized according to time.

In fact, a discrete point coverage is generally characterized by a finite domain consisting of a set of irregularly distributed points (ISO/FDIS 19123 2005). When these points can be arranged in a regular way, we may use Grid coverages. Another possibility to cover a continuous domain consists on partitioning the domain in a regular way in relation to the points of the discrete point coverage (i.e. tessellation). In this second case, we my use discrete surface coverages. This is the case for Swath and Radial scientific data types.

CDM profile and trajectory data types

Profile and Trajectory data are introduced as Point Observation sub-types. Thus, they may be consistently modeled as Discrete Point Coverage sub-types (see Fig. 12). However, in FES it is common to refer to profile and trajectory as the curves inferred from their observation points. Hence they may be modeled as Discrete Curve Coverage instances, as well. In this second case, which is semantically richer, the coverage finite domain consists of curves.

ISO 19123 provides a coverage sub-type called: Discrete-CurveCoverage. This class is a discrete coverage with the restriction that its domain consists of curves. Although the specification refers to its domain as "a finite spatial domain" (ISO/FDIS 19123 2005), the curve coverage domain is not limited to space, following the general coverage model previously discussed and, represented in Fig. 14. In fact, the DiscreteCurveCoverage has the restriction that the associated GeometryValuePairs shall be limited to CurveValuePairs (see Fig. 14). CurveValuePair is a subtype of GeometryValuePair that has a curve (i.e. the Curve type defined in ISO 19107) as the value of its geometry attribute. 19107:Curve is a geometry primitive subtype, defined by ISO 19107 (see magenta concepts of Fig. 14), and may be defined in any Euclidean space.

Table 1 Conceptual mapping for the CDM scientific data types

| CDM Scientific Data Type	ISO 19123 Coverage type
Gridded Data	
Unstructured	DiscretePointCoverage
Structured	DiscreteGridPointCoverage
Swath	DiscreteSurfaceCoverage
Point Observation Data	
Unconnected	DiscretePointCoverage
Station observation /Timeseries	DiscretePointCoverage
Trajectory	DiscretePointCoverage or DiscreteCurveCoverage
Profile	DiscretePointCoverage
Radial	DiscreteSurfaceCoverage or DiscreteCurveCoverage
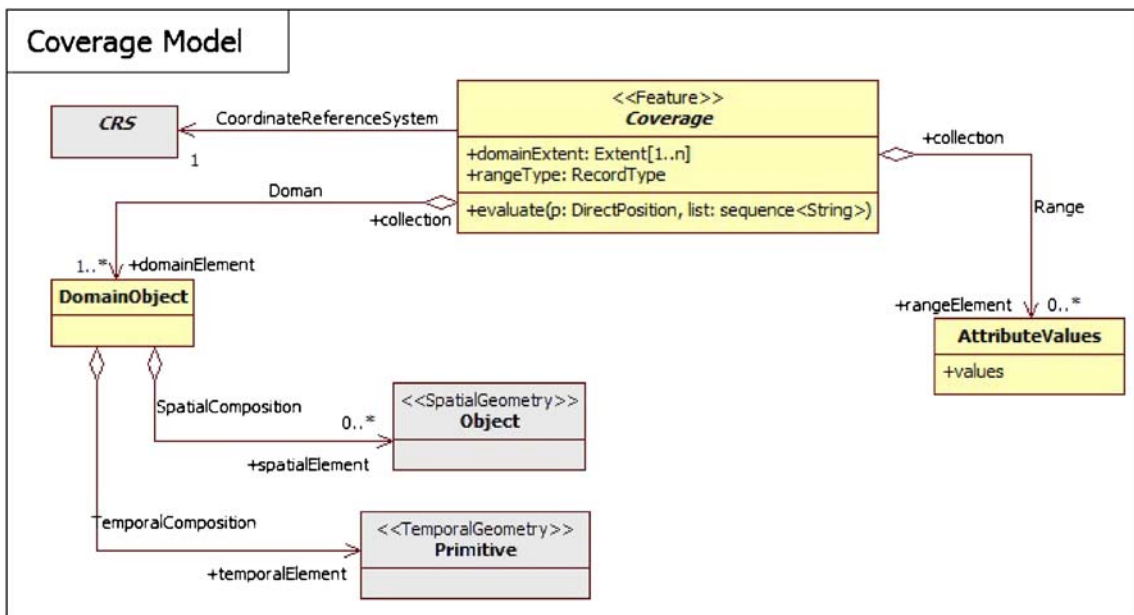
Fig. 13 ISO 19123 coverage abstract model

In fact, 19107:Curve is the basis for one-dimensional geometry: "a curve is a continuous image of an open interval and so could be written as a parameterized function such as $c(t)$: $(a, b) \rightarrow E^n$ where "t" is a real parameter and E^n is Euclidean space of dimension n (usually two or three, as determined by the coordinate reference system)" (ISO/IS 19107 2003).

CDM grid data types

Many dataset in the atmospheric and oceanic sciences contain gridded data to improve data storage and access. For example CF metadata conventions for netCDF are most well developed for the CDM Grid Scientific Data Type. In
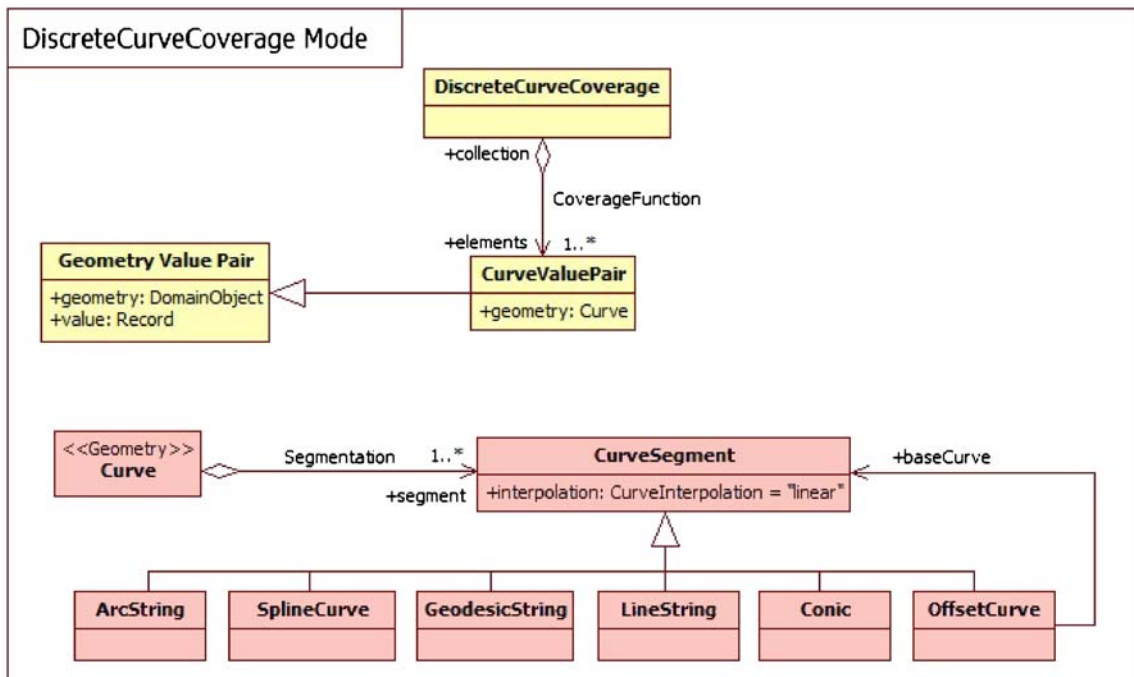
Fig. 14 ISO 19123 discrete curve model (simplified view)
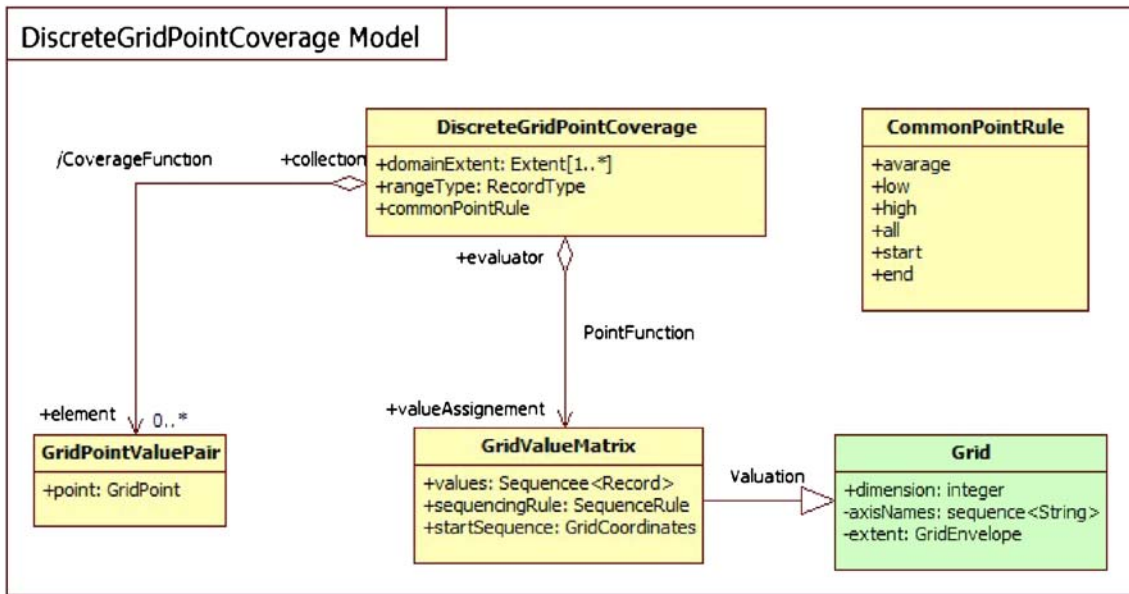
Fig. 15 Discrete grid point coverage model

fact, gridded data implements a systematic tessellation of the domain, employing a sequential enumeration of the elements of the domain. Generally, the tessellation represent how the data were acquired or how they were computed in a model.

In the realm of ISO data models, a grid coverage type is defined (ISO/FDIS 19123 2005). Grid is defined as a network composed of two or more sets of curves in which the members of each set intersect the members of the other

Fig. 16 Grid geometry model

sets in an algorithmic way. These curves partition a space into grid cells. The axes of the grid provide a basis for defining grid coordinates. The axes need to be identified to support sequencing rules for associating fea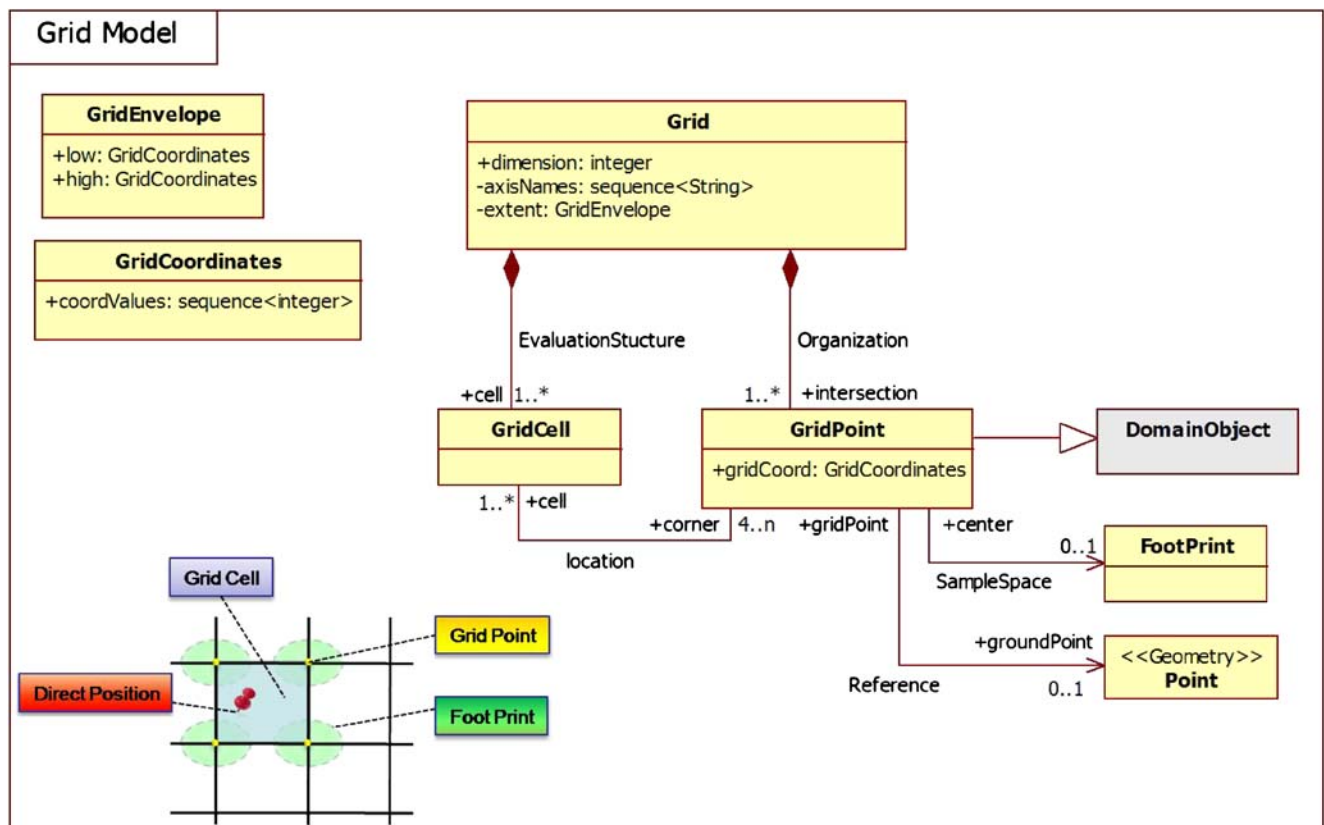ture attribute value records to the grid points. (ISO/FDIS 19123 2005). There are grid points at all grid line intersections; they represent the domain elements. Thus, FES gridded data may be effectively mapped onto Discrete Point Coverages whose domain consists of the point objects characterizing the grid tessellation. Therefore, as noted in the Table 1, for the CDM Grid data type the corresponding ISO coverage is the DiscreteGridPointCoverage.

The domain of a DiscreteGridPointCoverage instance is a set of GridPoints that are associated with records of feature attribute values through a GridValuesMatrix element. Certainly, DiscreteGridPointCoverage occurrences must be used to implement gridded-based coverage domains—either regularly or quasi-regularly spaced ones. Figures 15 and 16 depicts the DiscreteGridPointCoverage model and the related Grid model, respectively.

Referring to Figs. 15 and 16, the diagram references some important ISO elements (i.e. attributes, associations, classes, etc.) which are described in Table 2.

Discrete vs continuous coverage types

In most cases, a continuous coverage is also associated with a discrete coverage that provides a set of control values to be used as a basis for evaluating the continuous coverage (see Fig. 17). Evaluation of the continuous coverage at other direct positions is done by interpolating between the geometry value pairs of the control set. This often depends upon additional geomet- ric objects constructed from those in the control set; these additional objects are typically of higher topolog-ical dimension than the control objects (ISO/FDIS 19123 2005).

In ISO 19123, such objects are called "geometry value objects". A geometry value object is a geometric object associated with a set of geometry value pairs that provide the control for constructing the geometric object and for evaluating the coverage at direct positions within the geometric object.

A common example of geometry value object is represented by quadrilateral grid cell whose vertex are represented by four grid points (i.e. the set of geometry value pairs). The continuous quadrilateral grid coverage model is depicted in Fig. 17; the grid model is depicted in Fig. 16.

Table 2 ISO 19123 element specifications

| Element	Description
domainExtent	Contains the extent of the domain of the coverage; extents may be specified in space, time or space-time. The primitives to be used are defined in ISO 19107 (BADC Datasets: CF conventions:http://badc.nerc.ac.uk/help/formats/netcdf/index_cf.html) and 19108 (NetCDF: http://www.unidata.ucar.edu/software/netcdf/)
rangeType	Describes the range of the coverage. It consists of a list of attribute name/data type pairs. Both simple lists and recursive complex structures are possible, according to record data type defined in ISO/TS 19103
Coordinate Reference System (19111:CRS)	Links the Coverage to a unique coordinate reference system to which the objects in its domain are referenced. Hence, a coverage with the same range but with its domain defined in a different coordinate reference system is a different coverage. The Coordinate Reference System element is defined in ISO 19111 (OGC: Open Geospatial Consortium: http://www.opengeospatial.org/)
GridValueMatrix	Ties feature attribute values to grid geometry. It has three attributes: values, sequencingRule and startSequence. It holds a sequence of records associated with a sequencing rule that specifies an algorithm for assigning records of feature attribute values to grid points. An instance of the GridValuesMatrix may be either a generic Grid or one of its subclasses: RectifiedGrid (grid associated with an affine transformation between the grid coordinates and the coordinates of an external coordinate reference system; for example, a system related to the earth by a datum) and ReferenceableGrid (grid associated with a transformation that can be used to convert grid coordinate values to values of coordinates referenced to an external coordinate reference system; for example a system related to the earth by a datum).
PointFunction	links the DiscreteGridPointCoverage to the GridValuesMatrix for which it is an evaluator. The range of a coverage shall be a homogeneous collection of records: the range has a constant dimension over the entire domain, and each record field provides a value of the same attribute type over the entire domain
CoverageFunction	Links the discrete Coverage to the set of GridPointValuePair included in the coverage. This element may be generated from the GridValuesMatrix through the association PointFunction
GridPointValuePair	Describes an element of a set that defines the relationships of a discrete grid point coverage. It consists of two parts: a domain object from the domain of the coverage to which it belongs (i.e. a GridPoints geometry) and a record of feature attribute values from the range of the coverage to which it belongs (i.e. a feature attribute value Record)
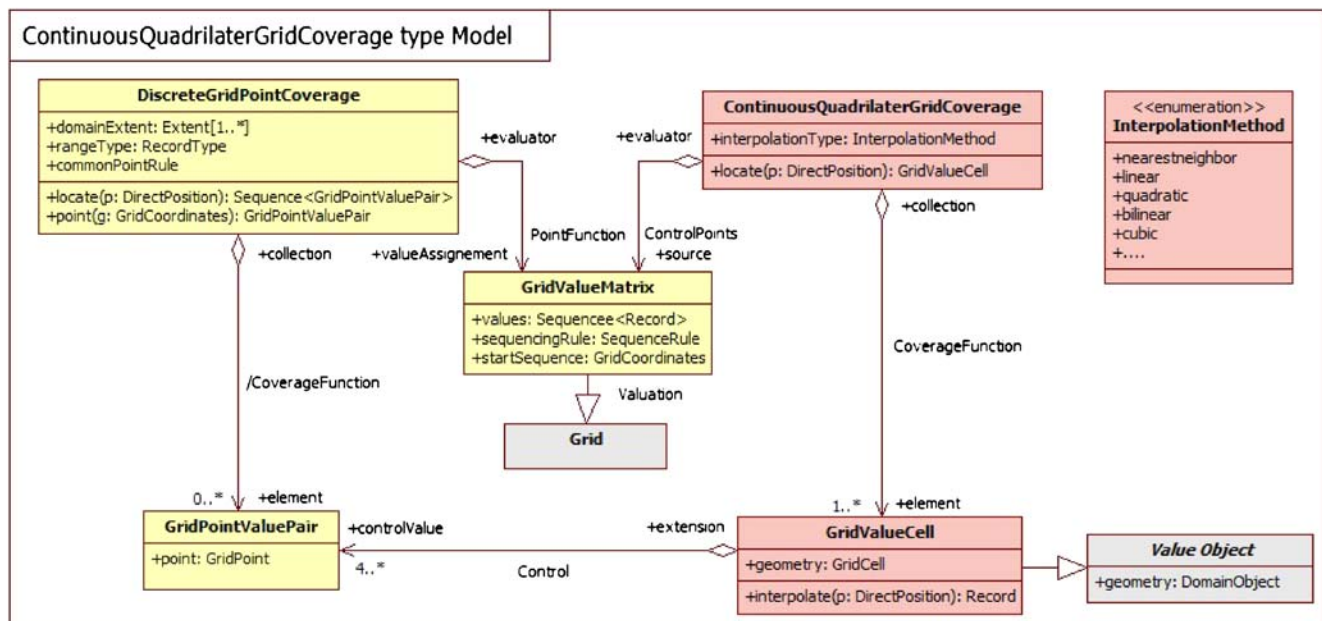
Fig. 17 Continuous quadrilater grid coverage model

In the FES domain, the continuous quadrilateral grid coverage type is associated to a discrete grid point coverage type by sharing the same geometry grid and matrix values; referring to Fig. 17, the two coverage subclasses share the GridValueMatrix object and the derived GridPointValuePair objects. The real difference consists in the realization of the *locate()* operation, which is inherited from the Coverage super-type. Therefore, "the principal use of discrete point coverages is to provide a basis for continuous coverage functions, where the evaluation of the continuous coverage function is accomplished by interpolation between the points of the discrete point coverage" (ISO/FDIS 19123 2005).

The evaluate operation for discrete grid point coverages

DiscreteGridPointCoverage is the Coverage subclass that returns the same record of values for any direct position within the sample space of a single grid point object in its domain (see Fig. 17 and 16). In fact, a grid point may be associated to a sample space: the footprint (see Fig. 16). The GridPoint is at the center of the sample space. The operation *evaluate* accepts a DirectPosition as input, locates the GridPointValuePairs that include the GridPoints containing the DirectPosition, and returns a set of values. The operation *evaluate* uses the GridValuesMatrix element to assign values to the GridPointsValuePairs. Normally, the input DirectPosition will fall within only one GridPointValuePair, and the operation will return the record of values associated with that GridPointValuePair. If the DirectPosition falls on the boundary between two GridPoint sample spaces, the operation will return a record of values

calculated according to the value of the *commonPointRule* attribute that characterizes the Coverage object.

The evaluate operation for continuous quadrilateral grid coverages

ContinuousQuadrilaterGridCoverage is the subclass of Coverage that returns a distinct record of values for any direct position within its domain.

The operation *evaluate* accepts a DirectPosition as input and returns a record of values for that direct position. The input DirectPosition will fall within one GridValueCell (i.e. a GridCell domain object) and the operation will return a record of values interpolated within that GridValueCell. If the DirectPosition falls on the boundary between two GridValueCell, the operation will return a record of values calculated according to the value of the *commonPointRule* attribute that characterizes the Coverage object.

In the case of FES data, the interpolation methods specified by the ISO continuous coverage classes do not apply in general. In fact, in most cases, any scientifically realistic interpolation depends on the physics of the situation as well as the geometry. Hence, any realistic interpolation is actually data dependent.

Therefore the CDM Gridded data types don't implement the evaluation operation using interpolation methods. They are mapped to the ISO discrete coverages because they actually represent sampled points in a continuous space where the intermediate values depend on the solution to physics-based equations that depend on the values of the range data.
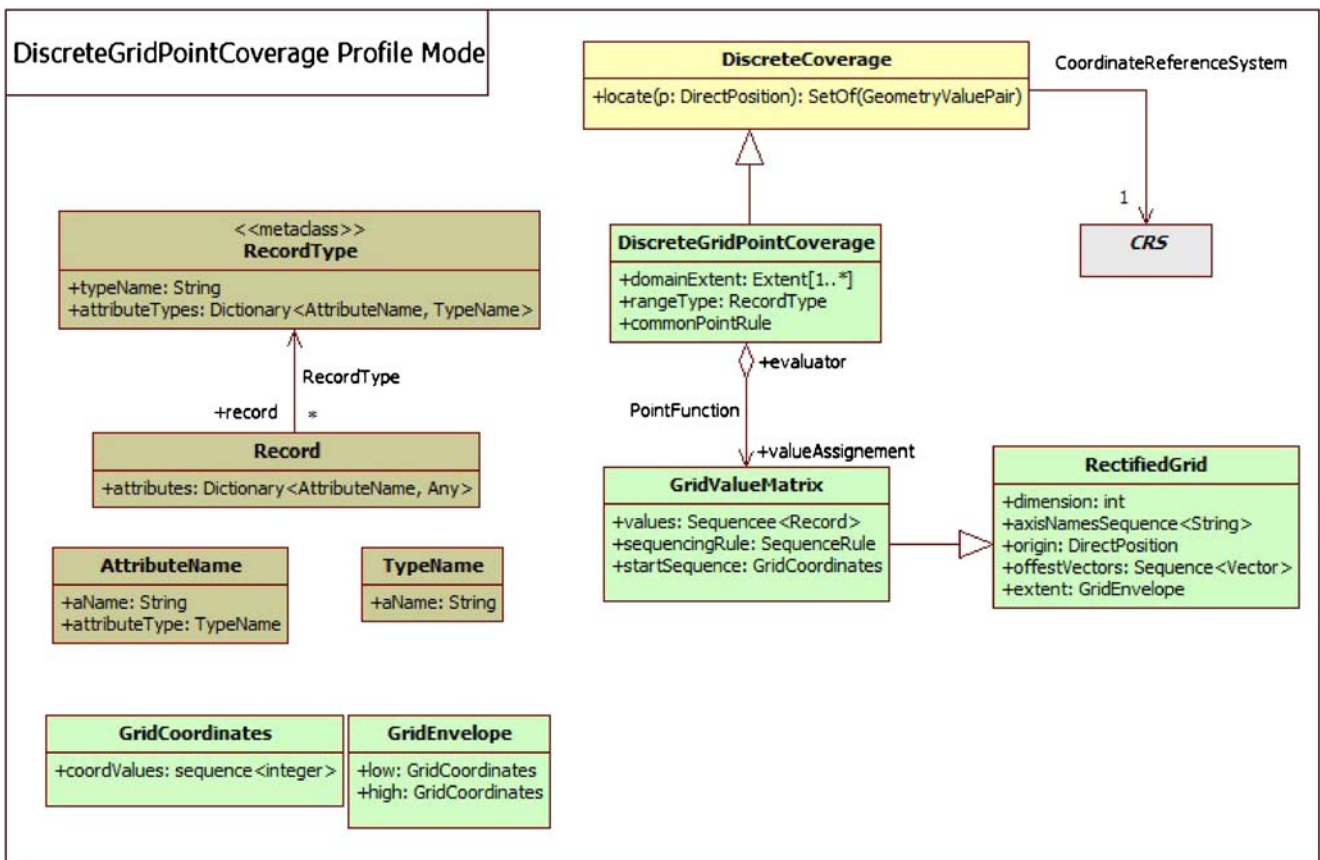
Fig. 18 Discrete grid point coverage profile model

Implementing the mapping from CDM grid data to ISO DiscreteGridPointCoverage

To explicitly map the CDM Grid data model (e.g. FES hyperspatial observation and model outputs) to the ISO Coverage data model (i.e. GIS coverage layers), there is a need to formalize the implicit knowledge, which characterizes FES dataset structuring and encodings, by using the ISO Coverage elements. This implies restructuring the FES data model, either introducing new, simplified data struc-

Fig. 19 CDM grid data type to ISO coverage mapping
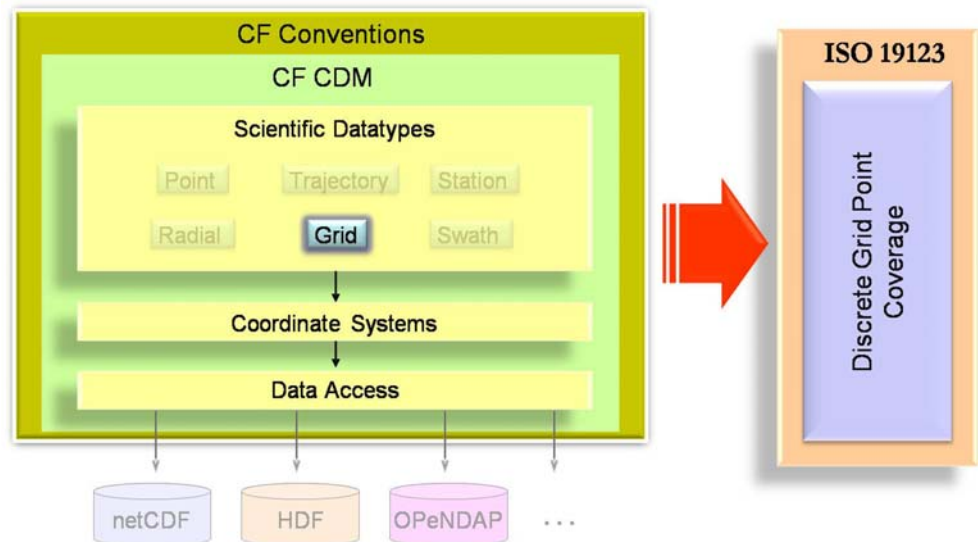
Fig. 20 CF-CDM grid data model profile

tures, or reinterpreting the existing concepts to foster general interoperability (e.g. interoperability with the GIS domain). Points to consider include:
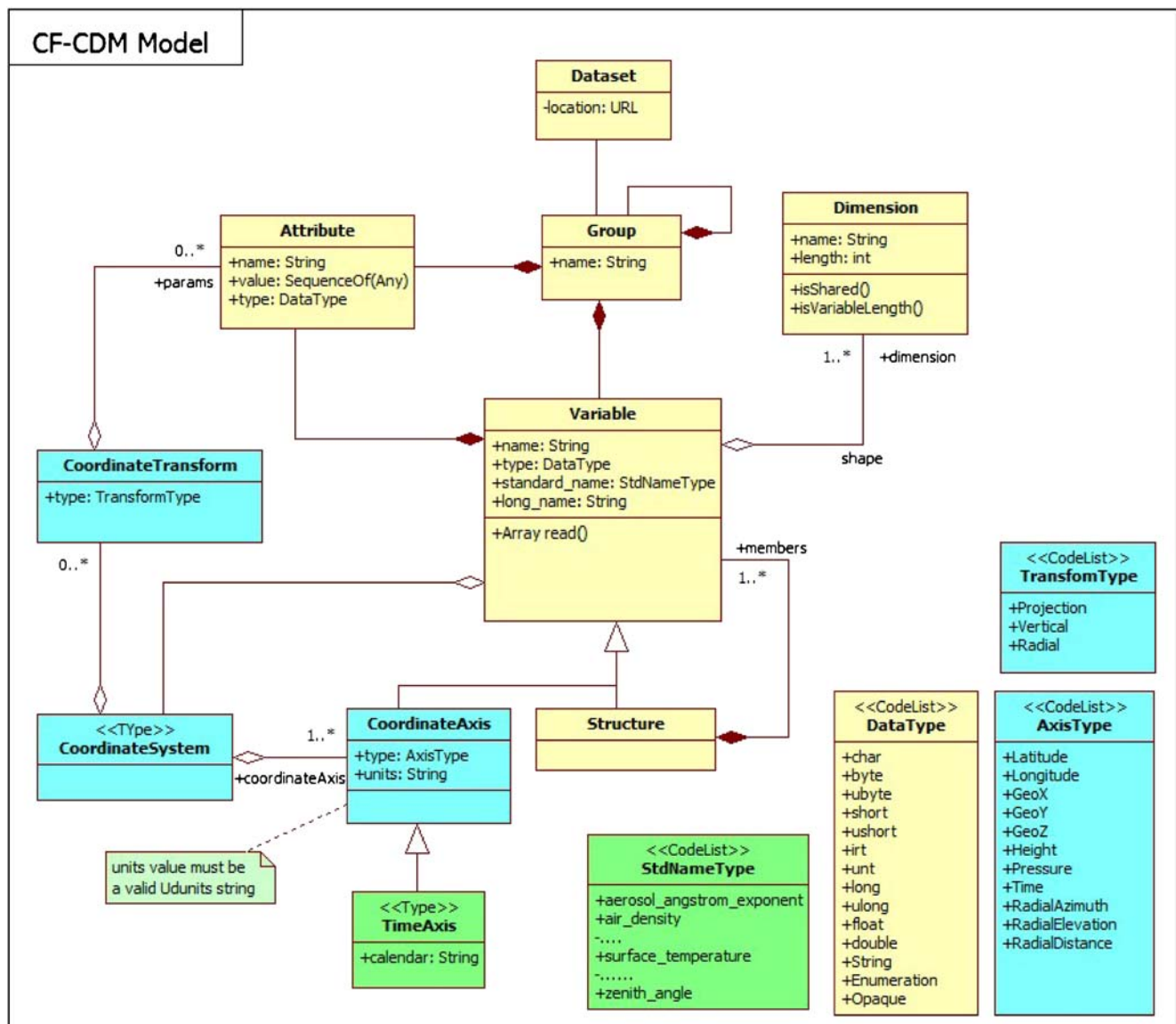
- CDM data model supports datasets characterized by multiple domains (e.g. more than one coordinate system is defined for a dataset), whereas an ISO coverage is characterized by a single coordinate system.
- CDM data model supports datasets characterized by arbitrary multi-dimensional domains, whereas an ISO coverage domain is either 2-D (space), 3-D (2D + vertical dimension), 4-D (2D + vertical dimension + time).
- Most commonly, CDM datasets grid axes coincide with reference system axes. However, CDM allows arbitrary domain shapes, i.e. grid axes ordering. Thus, it is possible to have a variable v1 defined on a grid <x, y, t, z> and a

variable v2 defined on a grid <z, x, t, y>. Since there is a fixed enumeration of allowed compound CRSs in ISO coverages, the transformation of such generalized grids coordinates to CRS coordinates may not be an affine transformation. In other words, mapping CDM grids to ISO (geo)rectified grids may require axes reshaping and reordering,

Therefore, the mapping must address these structural and semantics differences, applying the appropriate constraints and, hence, performing a complex mediation task.

There are two main steps to address these mediation issues: a first step consists in defining appropriate profiles for both CDM and ISO coverage data model, as far as grid point coverage is concerned. The second step deals with defining a set of mapping constraint rules.

The discrete grid point coverage profile

In keeping with the general ISO model for discrete grid point coverage, there exist several possible ways to describe and formalize the domain of discrete grid point coverages:

(1) Implementing a Grid object and its related Grid-PointValuePair objects (see Fig. 15): useful to model either regularly or quasi-regularly spaced domains.
(2) Implementing a Grid object (see Fig. 15), its associated GridPoint objects (see Fig. 16) and its valuation GridValueMatrix object: useful to model either regularly or quasi-regularly spaced domains.
(3) Implementing a RectifiedGrid object (see Fig. 18) and its valuation GridValueMatrix object: useful to model only regularly spaced domains.

As a starting point, we decided to model only regularly spaced domains, following the third solution. Therefore, a specific DiscreteGridPointCoverage profile was conceived. The profile model is shown in Fig. 18.

The CDM data model profile for generating discrete grid point coverages

In order to generate coverages from CDM grid datasets, it is important to recognize the minimum set of metadata elements which are mandatory to enable the mapping process. CF metadata conventions for netCDF data model (CF: Climate and Forecast Conventions: http://cf-pcmdi.llnl.gov/; CF Standard Name Table: http://cf-pcmdi.llnl.gov/documents/cf-standard-names/) are most well developed for the CDM Grid Scientific Data Type, providing most of the additional semantics required to achieve the mapping to ISO discrete grid point coverage model. Few other CF metadata entities (e.g. coordinate axis units) must be implemented by profiling the CDM data model, working out a CF-CDM profile. Figure 19 depicts the mapping general framework.

CF-CDM grid scientific data model profile

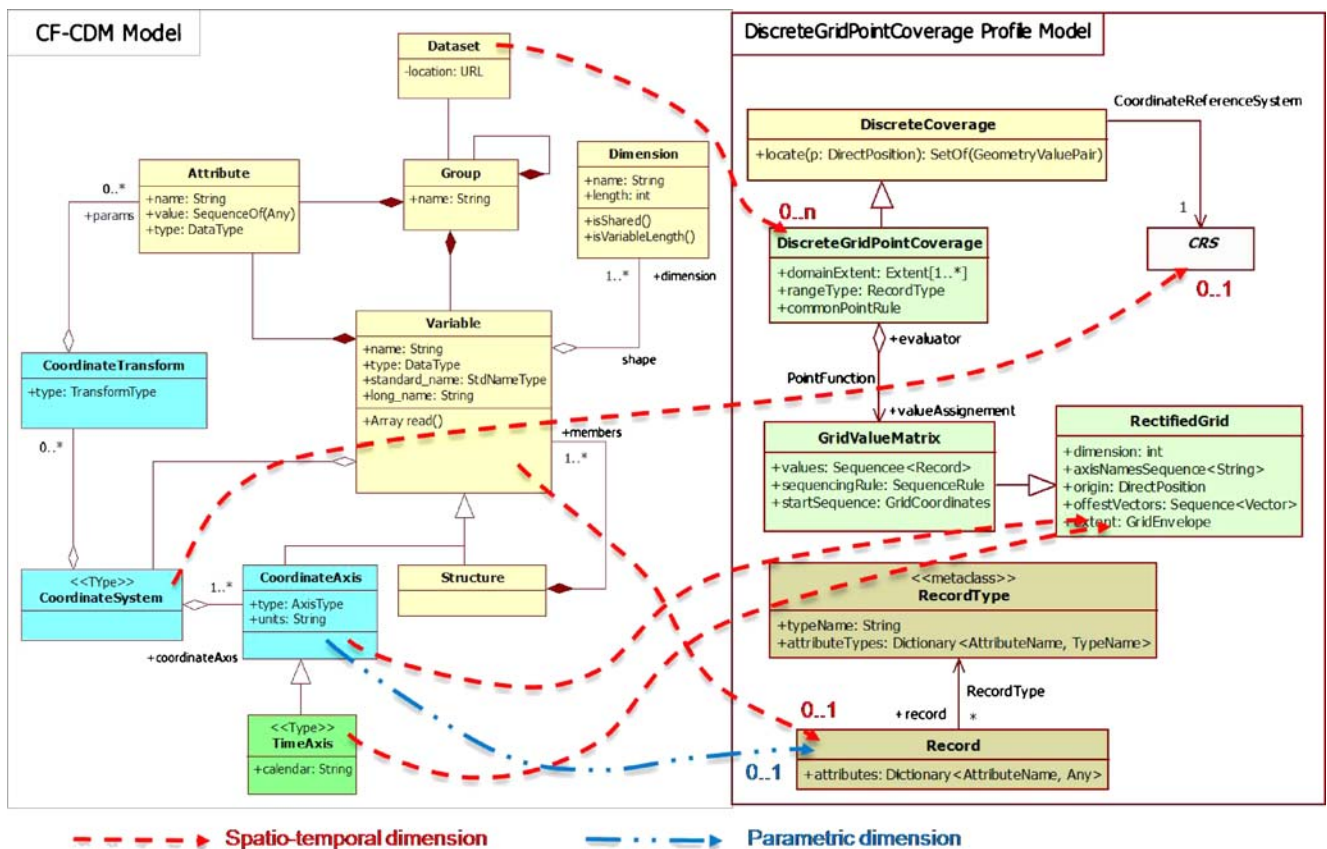Figure 20 shows the CF-CDM profile for mapping the CDM Grid data type.

Fig. 21 Mapping between the CF-CDM Grid and ISO DiscreteGridPointCoverage profile models

Table 3 CF-netCDF to ISO DiscreteGridPointCoverage mappings

CF-CDM grid data entity	*ISO 19123 DiscreteGridPointCoverage profile*
CF-Dataset	One or more DiscreteGridPointCoverage
CF-Variable	Record attribute
CoordinateSystem	CRS
CoordinateAxis allowed in CRS	CRS axis
CoordinateAxis not allowed in CRS (i.e. parametric dimension axis)	Record attribute
Range-set	List of Records

As shown in the picture, several CF convention features have been neglected at present. In particular, features used to accommodate projected CRS, as well as support to climatological statistics, cell boundaries, slanted/compressed grids or non-numeric coordinate axis will be further investigated in the future.

The data models mapping

Figure 21 depicts a high-level abstract mapping from the CF-CDM grid profile model to the DiscreteGridPointCoverage profile model. The dotted lines are intended to show the correspondence between concepts in the CF-CDM and in the DiscreteGridPointCoverage models. In particular, concepts in the CF-CDM model may have more than one counterpart concept, or none (as indicated by the usual multiplicity ranges on the arrow ends).

This high-level correspondences should provide the basis for the actual logical and physical mappings between realizations of the two abstract models. Actually, the proposed mapping is based on an experimental mapping between two such realizations of CF-CDM and ISO DiscreteGridPointCoverage: respectively, netCDF-CF and the OGC Coverage, as implemented in WCS. The mapping was defined and implemented in the framework of GALEON 1 and 2 Interoperability Experiment (Nativi et al. 2005).

However, the definition and implementation details of such logical and physical mappings are out of the scope of this work.

Mapping rules

A CF-CDM grid dataset may include more than one DiscreteGridPointCoverage, since it may contain (groups of) variables with different CoordinateSystems (e.g. Latitude–Longitude, Latitude–Longitude–Height).

In principle, grouping the variables defined in a dataset by their CoordinateSystem, a DiscreteGridPointCoverage may be defined for each group.

Actually, the association of groups and coverages may not be one-to-one, since the concept of coordinate system in CF-CDM is wider than CRS; in fact, parametric coordinate systems are allowed in CF-CDM but not in CRS[2]. A coordinate system is of type parametric if a physical or material property is used as a dimension (ISO 19111–2 Geographic information—Spatial referencing by coordinates—Part 2: Extension for parametric values); valuable examples are pressure in meteorology and density in oceanography. Hence, some of the obtained coverages may be further grouped together. It is also possible that a CoordinateSystem entity does not contain any axes allowed in coverage CRS (i.e. only parametric dimension axes); the associated variables would then originate no DiscreteGridPointCoverage instance.

In general, only spatial and temporal coordinates in a CF-CDM CoordinateSystem become part of a coverage CRS, whereas parametric dimension axes are mapped to compound range set components.

The domain of each DiscreteGridPointCoverage is obviously characterized by an implicit geometry, that is a regularly spaced grid. In general, the grid geometry of a DiscreteGridPointCoverage may be slanted, with respect to the CRS axes, by specifying appropriate offset vectors. However, the selected profile of CF-CDM only permits orthogonal grids, that is grids which are aligned with the CoordinateSystem axes (future evolutions of the CF-CDM profile may support slanted/compressed grids, or even non-numeric coordinate axis, e.g. by means of netCDF-CF features as AuxiliaryCoordinateVariable and coordinates attributes).

The domain of each coverage may be described by the extent of the related coordinate axis variables (if present).

The range-set of each DiscreteGridPointCoverage is a list of records with an attribute for every related CF-CDM variable and for every CoordinateAxis of the CoordinateSystem that is not allowed in CRS (i.e. parametric dimension axes).

Table 3 further illustrates the mapping from the CF-CDM grid data model to that of ISO DiscreteGridPointCoverage data model.

To exemplify the mapping, let's consider a fictitious instance of a CF-CDM dataset, containing variables v1, v2, v3, respectively defined on CoordinateSystem cs1={Lati-

---

[2] Future extension to ISO 19111 (see ISO/CD19111–2) may permit parametric CRS, that would accommodate the pressure axis.

tude, Longitude}, cs2={Latitude, Longitude, Height}, cs3= {Latitude, Longitude, Pressure}, as follows:

```
Dataset{
        CoordinateSystem cs1=<Latitude, Longitude>
        CoordinateSystem cs2=<Latitude, Longitude, Height>
        CoordinateSystem cs3=<Latitude, Longitude, Pressure>

        CoordinateAxis Latitude=<20, 22, 24>
        CoordinateAxis Longitude=<40, 41, 42>
        CoordinateAxis Height=<10, 11>
        CoordinateAxis Pressure=<100, 200>

        Variable v1=<cs1, <1, 2, 3, 4, 5, 6, 7, 8, 9>>
        Variable v2=<cs2, <1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
14, 15, 16, 17, 18>>
        Variable v3=<cs3, <1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
14, 15, 16, 17, 18>>
}
```

In the above syntax, a Variable is represented as an ordered list (delimited by the acute brackets '<' and '>') with a CoordinateSystem and an ordered list of (scalar) values. The association between a value and its coordinates in the associated CoordinateSystem is implied by its position in the list, according to the usual linearization of indices of multi-dimensional arrays (cfr. the C language). E.g. expressing v1 values as a bidimensional array $v1'[i][j]$, where i is the i-th Latitude value and j is the j-th Longitude value, then we have $v1'[i][j]=v1[i*|Longitude| + j]$. Hence, at 20°N, 41°E, v1=2.

When applied to v1 and v2, the mapping described above would include the following coverages (each list of records in the Rangeset contains a single record, so a few extra '{}' have been neglected):

```
Coverage{
        CRS=<Latitude, Longitude>
        Grid={origin=<20, 40>, offsetVectors=<<2, 0>, <0, 1>>
            }
        Rangeset=<{<v1, 1>}, {<v1, 2>}, {<v1, 3>}, {<v1, 4>},
{<v1, 5>}, {<v1, 6>}, {<v1, 7>}, {<v1, 8>}, {<v1, 9>}>
}
Coverage{
        CRS=<Latitude, Longitude, Height>
        Grid={origin=<20, 40, 10>, offsetVectors=<<2, 0, 0>, <0,
1, 0>, <0, 0, 1>>
            }
        Rangeset=<{<v2, 1>}, {<v2, 2>}, {<v2, 3>}, {<v2, 4>},
{<v2, 5>}, {<v2, 6>}, {<v2, 7>}, {<v2, 8>}, {<v2, 9>}, {<v2,
10>}, {<v2, 11>}, {<v2, 12>}, {<v2, 13>}, {<v2, 14>}, {<v2,
15>}, {<v2, 16>}, {<v2, 17>}, {<v2, 18>}>
```

When applied to v3, the mapping described above would discard the Pressure axis, which is not allowed in CRS, and originate the following coverage:

```
Coverage{>
        CRS=<Latitude, Longitude>
        Grid={origin=<20, 40>, offsetVectors=<<2, 0>, <0, 1>>
            }
        Rangeset=<{{<v3, 1>, <Pressure, 100>}, {<v3, 2>,
<Pressure, 200>}},
                {{<v3, 3>, <Pressure, 100>}, {<v3, 4>,
<Pressure, 200>}},
                …
                {{<v3, 17>, <Pressure, 100>}, {<v3, 18>,
<Pressure, 200>}}>
}
```

As shown above, the Rangeset is a list of tuple sets, where a tuple is an (unordered) set of couples <name, value>. The limits of the grid axes (not shown in this example) are derived by the originating dataset. The association between a tuple set and its coordinates in the coverage CRS is implied by its position in the list, as detailed above. Hence, at 20°N, 41°E, we have v3=3 at Pressure=100, v3=4 at Pressure=200.

Having the same CRS, the first and third coverage may be further grouped, as follows (this is a slight waste of space, with this encoding):

```
Coverage{
        CRS=<Latitude, Longitude>
        Grid={origin=<20, 40>, offsetVectors=<<2, 0>, <0, 1>>
            }
        Rangeset=<{{<v3, 1>, <Pressure, 100>, <v1, 1>}, {<v3,
2>, <Pressure, 200>, <v1, 1>}},
                {{<v3, 3>, <Pressure, 100>, <v1, 2>}, {<v3,
4>, <Pressure, 200>, <v1, 2>}},
                …
                {{<v3, 17>, <Pressure, 100>, <v1, 9>}, {<v3,
18>, <Pressure, 200>, <v1, 9>}}>
}
```

Lastly, since cs2 includes cs1, all the three coverages could in principle be grouped into one.

Conclusions and future work

As far as Earth Sciences (ES) are concerned, a unified data model called the Common Data Model (CDM) was

introduced. CDM unified data model implements a high level abstract model for accessing and using heterogeneous ES datasets. In fact, this model implements an abstract and unified interface to access well-accepted ES data models, such as: netCDF, HDF and GRIB.

In order to support the international effort on geo-spatial information interoperability, the CDM mapping into the corresponding elements of ISO 19123 coverage data model was presented at the abstract level. This mapping is important to facilitate the exploitation of ESS datasets for important societal benefit areas, such as: biodiversity, security and risk management, environmental policy and land management, policy for sustainable development etc.

The present CDM Scientific data types were mapped onto ISO coverage types. The case of CDM Grid data type content mapping was presented and discussed in detail. This mapping was achieved by profiling both the CDM Grid data model and the ISO discrete grid coverage model. The CF-CDM profile was introduced by applying the entire set of CF conventions. A specific implementation of ISO discrete grid coverage was selected for the mapping purpose. This mapping might provide a valid contribution to the specification of the OGC WCS 1.x profiles for specific coverage encoding formats, such as the CF-netCDF. In fact, a set of mapping rules, expressed in natural language, are introduced which should be applied as constraints for implementing the logical and physical mappings. Implementation may be realized using different technological frameworks, such as: XML schemas and XSLT, object-oriented classes and mapping operations. An example of the result obtained applying the proposed mapping rules is reported for a complex CDM Grid dataset.

Future work will consider:

• To conceive more detailed CF extensions for the other CDM scientific data types; then, implement the mappings between the extended CF and ISO coverage types for the other CDM scientific data types
• To introduce new interesting CDM scientific data types and consider their mappings to ISO coverage types
• To determine which protocol specifications are appropriate for providing access to the various scientific data

types. For example, one can envision serving these data via OGC Web Coverages Service (WCS), but also via OGC Web Feature Service (WFS) since a coverage is a special type of feature, after all. And finally, OGC Sensor Observations Service (SOS) might also come into play because many of these data originate on a variety of sensors
• To investigate the use of compound rangeset structures to improve the basic proposed mapping. In fact, the current mapping may not be correct in the case of compound parametric CRS –presently ISO 19111 does not support this type of coordinates but they are common in FES. Besides, this extension would improve performances

References

Booch G, Jacobson I, Rumbaugh J (1998) The unified modeling language user guide. Addison-Wesley, Reading, MA
European Commission (2006) "Communication from the Commission to the Council and the European Parliament: Interoperability for Pan-European eGovernment Services", COM(2006) 45 final, Brussels
IADBC (Interoperable Delivery of European eGovernment Services to public Administrations, Business and Citizens) Programme (2005) "Content Interoperability Strategy Working Paper, IADBC document
ISO/IS 19107 (2003) Geographic information—Spatial schema, ISO/IS 19107:2003(E)
ISO/IS 19108 (2002) Geographic information—Temporal schema, ISO 19108:2002(E)
ISO/IS 19111 (2003) Geographic information—Spatial referencing by coordinates, ISO 19111:2003(E)
ISO/FDIS 19123 (2005) Geographic information—Schema for coverage geometry and functions, ISO/FDIS 19123:2005(E)
Nativi S, Caron J, Davis E, Domenico B (2005) Design and implementation of netCDF Markup Language (NcML) and its GML-based extension (NcML-GML). Comput Geosci J 31 (9):1104–1118
W3C (2005) "Semantic Integration & Interoperability Using RDF and OWL", W3C Working Draft